SAND2017-8865C

Predictive Fidelity, Interpretability, and Resilience of Machine Learning Methods Applied to Scientific Simulations

Bert Debusschere, Ali Pinar, Khachik Sargsyan, Jeremy Templeton, Habib Najm
Sandia National Laboratories, Livermore, CA
bjdebus@sandia.gov

Machine Learning (ML) methods have shown tremendous potential for enabling knowledge gathering from extreme scale scientific simulations, such as extracting information from large data sets, discovering correlations in high-dimensional data, constructing Neural Network (NN) input-output maps as surrogate models in UQ operations, or learning/discovering the boundaries between qualitatively different behaviors in different regions of phase/configuration space. Despite the impressive capabilities developed in ML, however, prediction is only the first step. For scientific discovery, we need to be able to assess the *confidence* in the prediction, and to understand the *causes* behind this prediction. Further, for use in extreme scale computing and especially in the regime referred to as Beyond Moore's law computing, we need to get a good understanding of the *resilience* of ML algorithms with respect to system faults. Addressing these challenges will require a combination of expertise in machine learning, applied math, statistics, uncertainty quantification, and computer science, which makes it ideally suited for the DOE ASCR Applied Mathematics portfolio. This white paper outlines research questions towards a) assessing the confidence in, b) improving the interpretability of ML predictions, and c) enhancing the resilience of ML tools in extreme scale scientific computing.

Confidence: ML methods are often very accurate, but when they are not accurate, it is hard to know this ahead of time, which can lead to catastrophic failures. For scientific discovery, it is essential to be able to analyze and assess the confidence in ML predictions. This will allow us to assess the confidence in simulations that rely on ML tools, and also guide the allocation of more resources (more data or higher fidelity models) in the areas where it will have the most impact. Very recently, studies have introduced probabilistic concepts into ML [Ghahramani 2015] to provide a better understanding of what *learning* is and to provide more mathematical rigor into the field of ML. Many connections exist between commonly used ML methods and Bayesian statistics, leading to a new field of research, commonly referred to as *Bayesian Deep Learning* (BDL) [Wang, 2016]. For example, a stochastic regularization technique called *dropout* in ML to reduce overfitting has strong analogies with Bayesian evidence maximization [Blundell, 2015; Gal, 2015] and can be used to assess model uncertainty in NNs and other deep learning models [Gal, 2015b].

Besides a better mathematical understanding, one of the outcomes of these new methods is that they allow the assessment of the accuracy in classification problems and an analysis of the uncertainty in NN regression problems based on the availability of data [Blundell, 2015; Gal, 2015]. Many questions remain, however, in terms of the practical application of these approaches to large scale data sets, and how they compare to cross-validation and ensemble approaches. Some studies argue in favor of using variational Bayesian methods to compute approximate distributions of the NN model coefficients that capture the uncertainty [Blundell, 2015]. Other studies maintain that while Bayesian statistics provide a good mathematical interpretation, it is more practical to stick with dropout from a computational perspective [Gal, 2015].

While the studies above are promising leads towards quantifying the confidence in ML predictions, they are only scratching the surface, and a much stronger mathematical underpinning of the ML tools is required. For example, what is the convergence rate as a function of the amount and nature of the training data? How does this depend on the dimensionality of the data space? When can we expect convergence? What is the best way to select metaparameters? How

do we handle aleatory and epistemic uncertainties? How can experimental design be coupled with ML?

Interpretability: Interpretability, also referred to as *explainability,* pertains to the question of why ML methods make predictions in a certain way. For example, a Deep Neural Network (DNN) automatically selects the set of features that are most useful for a specific classification problem. Why do these features specifically serve well for the classification of the phenomena of interest? Answering this question gives insight into the underlying dynamics of the system being studied. For simple classifiers such as support vector machines, or decision trees, this question can be tractable. However, for more complex classifiers such as DNNs or random forests, answering that question is non-trivial, and requires novel approaches to improve their interpretability [Hara, 2016]. Without insights into decision mechanisms of methods like DNNs and Random forests, we have to treat the classifier as a black box, and build an interpretable classifier around a point of interest by generating new points and using the black-box classifier as the label generator, (see e.g., LIME). Building this classifier at the same time, especially for high dimensional problems remains as a challenge. Sampling needs to be done judiciously to preserve locality, yet identify the boundaries of classification. Moreover, locality of the feature space needs to be tied to the interpretable parameters of the simulation. Features that have predictive power (e.g., spectra of the matrices that govern the simulation) are not always interpretable at the application level.

Resilience: Last but not least, the need for methods that are resilient to system faults will become stronger as computing hardware pushes the extreme scale envelope, and even more so, for the hardware that is envisioned for the Beyond Moore's law computing era. Both neuromorphic and quantum computing hardware have a lot of randomness and variability in their operation, which will require an embrace of probabilistic approaches to characterize the reliability and repeatability of their results. For example, while NN based approaches in neuromorphic computing tend to be quite robust to system faults [Burr, 2017; Schuman, 2017], what are the limits of this resilience and what is the cost trade-off? What level of redundancy in terms of extra neurons and connections are needed to mitigate the impact of system faults? What are the implications of this required resilience on the prediction accuracy, device cost, power consumption, trainability, and speed of operation?

In conclusion, the emergence of powerful ML approaches offers tremendous opportunities for enabling and analyzing extreme scale scientific simulations. However, a stronger mathematical footing will be required to fully develop these opportunities, leading to many open research directions.

Acknowledgement:

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

References:

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight Uncertainty in Neural Networks. In *Proceedings of the 32nd International Conference on Machine Learning* (Vol. 37, pp. 1613–1622). Retrieved from http://arxiv.org/abs/1505.05424

Burr, G. W., Shelby, R. M., Sebastian, A., Kim, S., Kim, S., Sidler, S., … Leblebici, Y. (2017). Neuromorphic computing using non-volatile memory. *Advances in Physics: X*, *2*(1), 89–124. http://doi.org/10.1080/23746149.2016.1259585

Gal, Y., & Ghahramani, Z. (2015). Dropout as a Bayesian Approximation : Representing Model Uncertainty in Deep Learning. *ICML*, *48*, 1–10.

Gal, Y., & Ghahramani, Z. (2015b). On Modern Deep Learning and Variational Inference. In *Advances in Approximate Bayesian Inference workshop, NIPS* (pp. 1–9).

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459. http://doi.org/10.1038/nature14541

Hara, S., & Hayashi, K. (2016). Making Tree Ensembles Interpretable. *2016 ICML Workshop on Human Interpretability in Machine Learning*.

Riberio, M. T., Singh S., & Guestrin C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pages 1135–1144, New York, NY, USA.

Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., & Plank, J. S. (2017). A Survey of Neuromorphic Computing and Neural Networks in Hardware. *arXiv*, 1–88. Retrieved from http://arxiv.org/abs/1705.06963

Wang, H., & Yeung, D.-Y. (2016). Towards Bayesian Deep Learning: A Framework and Some Existing Methods. *IEEE Transactions on Knowledge and Data Engineering*, *28*(12), 3395–3408. http://doi.org/10.1109/TKDE.2016.2606428